2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) | 978-1-6654-0382-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ECTI-CON51831.2021.9454740

Reinforcement Learning-Based Adaptive Modulation for Vehicular Communication

Nazmia Kurniawati
Department of Electrical Engineering
Universitas Trisakti,
Jakarta, Indonesia
nazmia.kurniawati@trisakti.ac.id

Yuli Kurnia Ningsih
Department of Electrical Engineering
Universitas Trisakti,
Jakarta, Indonesia
yuli_kn@trisakti.ac.id

***Abstract*—Vehicular communications allowed the vehicles connected to traffic lights, road signs, or building infrastructures. Adaptive modulation enabled the system to maintain the quality level as the vehicles moved through different environments. In this paper, QPSK, 8PSK, and 16-QAM modulation schemes are utilized for low, medium, and high noise environments with the AWGN channel model. Reinforcement learning is implemented in the V2I scheme with an epsilon-greedy algorithm to decide which modulation scheme should be used based on the environment condition. The simulation showed promising results with most selected modulation scheme in low noise environment is 16-QAM, in medium noise is 8PSK, and in high noise is QPSK.**

***Index Terms*—Epsilon Greedy, Modulation, Reinforcement learning, V2I.**

I. INTRODUCTION

Vehicular communications is a technology that enables vehicles to be connected to other vehicles or infrastructures while on the move [1]. Vehicular communications include Vehicle-to-Vehicle (V2V) technology, Vehicle-to-Infrastructure (V2I), and Vehicle-to-Everything (V2X) [2]. As the vehicles moved, the channel conditions also changed. It resulted in different permitted Signal to Noise Ratio (SNR) to maintain a certain quality level.

Adaptive modulation enables the system to choose the best modulation scheme depends on the channel conditions [3]. A higher order modulation can be utilized when the channel is in good condition, as there are fewer noise sources exist compared to higher noise level environments. On the contrary, when the channel has a high noise level, the data is sent in lower data rates to avoid high packet loss. By implementing an adaptive modulation scheme, the bandwidth usage can be increased and sensitivity to environmental changes can be decreased [4]. Therefore, a certain quality level can be maintained while the vehicle moves through different environments.

Reinforcement learning is a part of machine learning where the learning process happens when the agent interacts with the environment [5]. Unlike machine learning that needed a dataset for training, an agent in reinforcement learning explored the environment and made the decision based on the defined reward and punishment value. After some trial and error actions, the agent learned what action gave the most optimum result [6]. So, the agent tended always to take the action that gives the maximum result.

Epsilon-greedy is one of the algorithms used for the learning process. It is an algorithm that balances exploration and exploitation based on epsilon value [7]. When exploring the environment, the agent randomly took action regardless of the received rewards. While during exploitation, the agent took the action that gives the highest rewards. Epsilon is the value that decides whether the agent explores or exploits.

In V2I technology, the vehicles are connected to surrounding infrastructures such as traffic lights, road signs, or building infrastructures. These infrastructures served as a transmitter or Tx, while the moving vehicle served as a receiver or Rx. Reinforcement learning with the epsilon-greedy algorithm acted as a decision-maker determining which modulation scheme should be used based on the environmental conditions. There are three modulation schemes used for different conditions: Quadrature Phase Shift Keying (QPSK), 8PSK, and 16 Quadrature Amplitude Modulation (16-QAM). The additive White Gaussian Noise (AWGN) channel model is used to distinguished the environment conditions.

The organization of this paper is as follows. The first section mentioned the general idea of this paper. Section two explained the algorithm implementation. The next section discussed the obtained result. Finally, the last section concluded the paper.

II. ALGORITHM IMPLEMENTATION

In this research, the vehicle is assumed to move through various noise-level environments with the AWGN channel model. The maximum allowed Bit Error Rate (BER) is set at the value 10^{-3}[8]. This value is used as the threshold for error probability (Pb). The modulation schemes used for adaptive modulation are QPSK, 8PSK, and 16-QAM. A modulation scheme with a higher data rate is considered for lower noise channel conditions, while the modulation with lower data is used for a higher noise environment.

The parameter such as Energy per Bit to the Spectral Noise Density (Eb/No) level [9] is used as an indicator to distinguish each environment. To calculate Eb/No, equation 1 is used to change Q function into error function (erfc) [10].

$$Q(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right) \tag{1}$$

Based on [10], equation 2 is used to calculate error probability of QPSK.

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \tag{2}$$

Then by using equation 1 to modify equation 2, the equation to calculate QPSK error probability becomes:

978-1-6654-0382-5/21/$31.00 ©2021 IEEE

$$P_b = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right) \qquad (3)$$

Equation 4 is used to calculate 8PSK error probability[10].

$$P_b = \frac{1}{log_2 M} 2Q\left(\sqrt{2\frac{E_b}{N_0} log_2 M} \;\; sin\frac{\pi}{M}\right) \qquad (4)$$

By modifying equation 4 with equation 1, then it becomes:

$$P_b = \frac{1}{3} erfc\left(\sqrt{3\frac{E_b}{N_0}} \; sin\left(\frac{180}{8}\right)^{o}\right) \qquad (5)$$

For 16-QAM, equation 6 is used to calculate the error probability.

$$P_b = \frac{4}{log_2 M}\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3 log_2 M}{M-1}\frac{E_b}{N_0}}\right) \qquad (6)$$

By substituting the Q function, the error probability equation becomes:

$$P_b = \frac{3}{8} erfc\left(\sqrt{\frac{2}{5}\frac{E_b}{N_0}}\right) \qquad (7)$$

Using equations 3, 5, and 7 with Pb 0.001, Eb/No is calculated. Table 1 showed Eb/No value obtained from the calculation.

TABLE I Eb/No THRESHOLD VALUE FOR MODULATION

Modulation	E_b/N_o (dB)
QPSK	6.78
8PSK	10
16-QAM	10.52

As shown in Table 1, the higher the modulation scheme, the higher the required Eb/No. With assumption the transmit power level from Tx is the same for all modulation schemes, then the noise level must be lower for the modulation scheme to achieve the required Eb/No with Pb 0.001.

When Eb/No is less than 6.78 dB, the environment is considered a high noise condition. Because with the constant transmit power, it needs a higher noise level to make Eb/No lower. When the Eb/No is between 6.78 and 10 dB, it is considered medium noise. Furthermore, if the Eb/No is more than 10 dB, it is considered low noise. Table 2 summarises the channel conditions.

TABLE II Eb/No VALUE FOR CHANNEL CONDITION

Noise	Value (dB)
High	$E_b/N_0 < 6.78$
Medium	$6.78 \leq E_b/N_0 \leq 10$
Low	$E_b/N_0 > 10$

After defining the environment's condition, each environment modulation scheme is classified. Considering the Eb/No threshold value, when it is less than 6.78 dB, the only possible modulation to be used is QPSK because the Pb will be more than 0.001 if 8PSK or 16-QAM is used. While the Eb/No is less than 10 dB, QPSK and 8PSK can be used. Moreover, when the Eb/No is more than 10 dB, all three modulation schemes are permitted because the Eb/No is higher than the defined threshold. Table 3 showed the mapping for each modulation scheme in different environmental conditions.

TABLE III MODULATION SCHEMES MAPPING FOR ENVIRONMENTS

		Modulation Schemes		
		16-QAM	8PSK	QPSK
Noise	High	X	X	V
	Medium	X	V	V
	Low	V	V	V

Note:
X : the modulation scheme isn't permitted to be used
V : the modulation scheme is permitted to be used

More than one modulation scheme is permitted in a high and a medium noise environment. Priority is used to control the selection of modulation schemes. A modulation scheme with a higher data rate is given a higher priority. Therefore the agent will tend to choose the higher modulation scheme.

During the observation, the agent will act in a random way. Since reinforcement learning is based on giving reward and punishment to the agent to obtain a learning experience, the punishment is given when the agent chooses the prohibited modulation scheme. In contrast, a reward will be given every time the agent chooses the permitted scheme. The reward value will be given according to modulation priority. The higher the priority, the higher the reward value. Table 4 showed the reward and punishment value for each condition.

TABLE IV REWARD AND PUNISHMENT VALUE

		Modulation Schemes		
		16-QAM	8PSK	QPSK
Noise	High	-1	-1	0.9
	Medium	-1	0.9	0.3
	Low	0.9	0.3	0.1

After the conditions and parameters are defined, the algorithm is built. The algorithm works in an episodic manner to simulate a vehicle's condition moving through several different environments. The program is looped 1000 times, illustrating 1000 environments that the vehicle passed in one trip.

At first, the transmitter randomly sets the modulation scheme. Then the communication between transmitter and receiver is established through the AWGN channel. The algorithm randomly set the channel conditions, either in low, medium, or high noise. The receiver then calculated the probability error value based on the modulation scheme. If the error probability is less than 10^{-3}, the reward is given. However, if the calculated error probability is more than 0.001, the agent is given a punishment. Using the epsilon-greedy algorithm, the agent learned what action should be taken to avoid punishment and gain a high reward; it is either command the transmitter to change the modulation scheme or use the current scheme. After that, the agent sends a command to the transmitter containing the decision. Then the modulation is set based on that order. Figure 1 showed the designed algorithm.

For the decision-making process, epsilon-greedy is utilized. Epsilon value controls the decision-making process. Random number (r) is generated randomly in order that the algorithm can switch between exploration and exploitation activities. During exploitation, the rewards and punishment value is taken from table 4. For simulation, a different epsilon value is utilized to see its effect on the agent's action. The epsilon values used for the simulation are 0.1, 0.2, 0.3, 0.4, and 0.5.

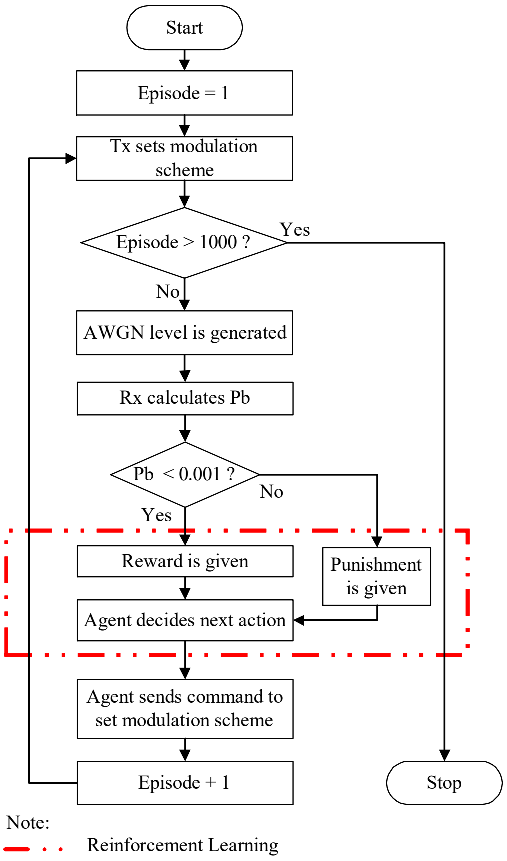

Fig 1. Algorithm Flow

III. RESULT

Figure 2 showed the simulation result of actions taken by the agent. During the 1000 episodes for each epsilon value, the agent took various actions that resulted in various rewards and even punishment. When epsilon is 0.1, the agent took over 900 actions having the highest rewards and took very few actions with low rewards. As the epsilon increased, the highest reward taken actions decreased while the lower rewards and punished actions increased. It happened because as the epsilon value increased, the gap between exploration and exploitation became wider. As the epsilon is getting higher, the agent's possibility to take random action instead of choosing the best action is also more significant.

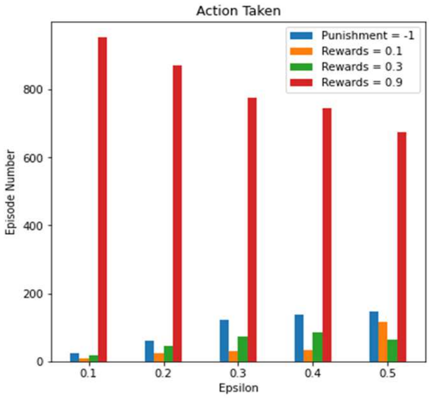

Fig 2. Rewards and punishment result

From the data shown in figure 2, the modulation scheme election is then broken down. Figure 3 displayed the modulation scheme distribution when the agent is in a low noise environment. As it can be seen, most of the time, the agent chose the 16-QAM scheme. It is as desired because 16-QAM has the highest priority among other schemes. As the epsilon value increased, 16-QAM became less preferred, and other lower bitrate modulations are chosen. The result has similar characteristics as shown in the figure before.

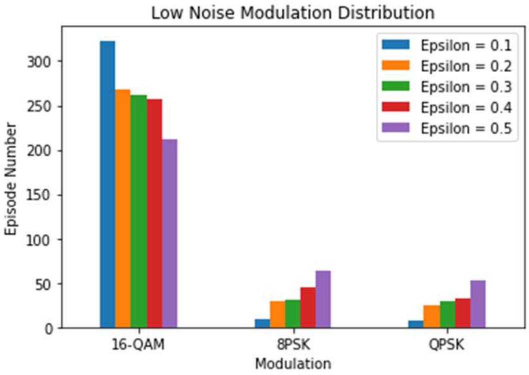

Fig 3. Modulation distribution in a low noise environment

Figure 4 showed the agent's modulation scheme distribution in a medium noise environment. It showed a similar result with modulation scheme distribution in a low noise environment. On most occasions, the agent chose 8PSK as the modulation scheme. As the epsilon value is getting higher, 8PSK selection is becoming lesser, and QPSK selection becomes preferable for the agent. Even though the 16-QAM selection is prohibited, the agent still chose the scheme caused by random action selection when the agent is in exploration mode. When implemented in the real world, it could lead to high packet loss or connection drop.

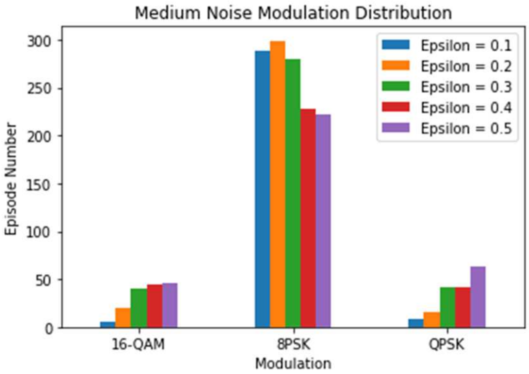

Fig 4. Modulation distribution in a medium noise environment

Figure 5 illustrated modulation scheme distribution when the agent is in a high noise environment. As illustrated by the figure, the QPSK modulation selection is getting lower as the epsilon value increases. In a high noise environment, 16-QAM and 8PSK modulation schemes should not be selected because the error probability will be bigger than 0.001. Those two modulation schemes are selected because the agent is in exploration mode when the generated random number (r) is less than the epsilon value. 16-QAM and 8PSK modulation scheme selection became more frequent as the epsilon value is getting higher because the agent's probability is in exploration mode is bigger. Therefore, modulation schemes are randomly chosen.

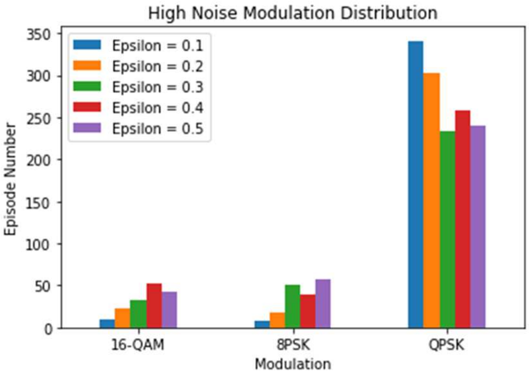

Fig 5. Modulation distribution in a high noise environment

IV. CONCLUSION

Reinforcement learning implementation with an epsilon-greedy algorithm is possible for V2I technology. The agent could choose the modulation scheme with the highest reward from simulations with 1000 different environments with epsilon value varied from 0.1-0.5. The simulation showed promising results with most selected modulation scheme in low noise environment is 16-QAM, in medium noise is 8PSK, and in high noise is QPSK. In some episodes, the agent took prohibited actions. It happened because of the algorithm's nature that chose a random modulation scheme during exploration mode. In order to improve the performance, a more sophisticated algorithm is required. Temporal Difference or Artificial Neural Network could be a promising candidate for that purpose since they have better computing ability.

REFERENCES

[1] F. Hu, *Vehicle-to-Vehicle and Vehicle-to-Infrastructure Communications*. New York: Taylor & Francis, 2018.

[2] F. Arena and G. Pau, “An Overview of Vehicular Communications,” *Futur. Internet*, vol. 11, no. 2, p. 27, Jan. 2019, doi: 10.3390/fi11020027.

[3] R. F. Masood, “Adaptive Modulation (QPSK, QAM),” *arXiv Prepr. arXiv1302.7145*, Feb. 2013, [Online]. Available: http://arxiv.org/abs/1302.7145.

[4] A. Novfitri, T. Suryani, and Suwadi, “Performance Analysis of Vehicle-to-Vehicle Communication with Adaptive Modulation,” in *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, Oct. 2018, pp. 187–191, doi: 10.1109/EECCIS.2018.8692895.

[5] S. Ravinchandiran, *Hands-On Reinforcement Learning with Python*. Birmingham: Packt Publishing Ltd., 2018.

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. London: The MIT Press, 2015.

[7] A. dos Santos Mignon and R. L. de Azevedo da Rocha, “An Adaptive Implementation of ε-Greedy in Reinforcement Learning,” *Procedia Comput. Sci.*, vol. 109, pp. 1146–1151, 2017, doi: 10.1016/j.procs.2017.05.431.

[8] A. Sassi, F. Charfi, L. Kamoun, Y. Elhillali, and A. Rivenq, “OFDM Transmission Performance Evaluation in V2X Communication,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 141–148, Oct. 2012, [Online]. Available: http://arxiv.org/abs/1410.8039.

[9] J. Pearce, “What’s All This Eb/No Stuff, Anyway?,” *Fall 2000 issue Spread Spectr. Scene Online*, vol. 7, no. 1, 2000.

[10] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: Mc. Graw-Hill, 2008.